

**WHAT IS CLAIMED IS:**

1 1. A method for detecting navigation bars in a document,  
2 the method comprising:  
3 a) segmenting the document into components; and  
4 b) for each of the components, determining whether or  
5 not the component is anchor-heavy, wherein if the  
6 component is anchor-heavy, it is determined to be a  
7 navigation bar.

1 2. The method of claim 1 wherein the act of determining  
2 whether or not the component is anchor-heavy is based on a  
3 number of anchors in the component and a number of  
4 non-anchor words in the component.

1 3. The method of claim 1 wherein the act of determining  
2 whether or not the component is anchor-heavy includes  
3 i) determining a number of anchors in the  
4 component,  
5 ii) determining a number of non-anchor words in  
6 the component, and  
7 iii) if the number of anchors is greater than a  
8 predetermined threshold and if the number of  
9 anchors is greater than the number of non-anchor  
10 words, then determining that the component is  
11 anchor-heavy.

1 4. The method of claim 3 wherein the predetermined  
2 threshold is about three.

1 5. The method of claim 3 wherein the predetermined  
2 threshold is three.

1 6. The method of claim 1 wherein the act of determining  
 2 whether or not the component is anchor-heavy includes  
 3 i) determining a first count to be a number of  
 4 anchors in the component,  
 5 ii) determining a second count to be a number of  
 6 non-anchor words in the component,  
 7 iii) incrementing the second count by the number  
 8 of words in an anchor having more words than a  
 9 predetermined threshold to determine a non-anchor  
 10 word count, and  
 11 iv) if the first count is greater than a second  
 12 predetermined threshold and if the first count is  
 13 greater than the non-anchor word count, then  
 14 determining that the component is anchor-heavy.

1 7. The method of claim 6 wherein the predetermined  
 2 threshold is about four.

1 8. The method of claim 6 wherein the predetermined  
 2 threshold is four.

1 9. The method of claim 1 wherein the act of segmenting the  
 2 document into components includes generating a parse tree  
 3 based on the document, wherein a first node corresponding  
 4 to a first component is a child of a second node of a  
 5 second component if the first component is included in the  
 6 second component.

1 10. The method of claim 9 wherein the act of determining  
 2 whether or not the component is anchor-heavy is based on  
 3 (i) a number of anchors in a node corresponding to the

4 component and all descendant nodes of the node, and (ii) a  
5 number of non-anchor words in the node corresponding to the  
6 component and all the descendant nodes of the node.

1 11. The method of claim 9 wherein the act of determining  
2 whether or not the component is anchor-heavy includes  
3 i) determining a number of anchors in a node  
4 corresponding to the component and all descendant  
5 nodes of the node,  
6 ii) determining a number of non-anchor words in  
7 the node corresponding to the component and all  
8 the descendant nodes of the node, and  
9 iii) if the number of anchors is greater than a  
10 predetermined threshold and if the number of  
11 anchors is greater than the number of non-anchor  
12 words, then determining that the component is  
13 anchor-heavy.

1 12. The method of claim 11 wherein the predetermined  
2 threshold is about three.

1 13. The method of claim 11 wherein the predetermined  
2 threshold is three.

1 14. The method of claim 9 wherein the act of determining  
2 whether or not the component is anchor-heavy includes  
3 i) determining a first count to be a number of  
4 anchors in a node corresponding to the component  
5 and all descendant nodes of the node,  
6 ii) determining a second count to be a number of  
7 non-anchor words in a node corresponding to the  
8 component and all descendant nodes of the node,

9           iii) incrementing the second count by the number  
10           of words in an anchor having more words than a  
11           predetermined threshold to determine a non-anchor  
12           word count, and  
13           iv) if the first count is greater than a second  
14           predetermined threshold and if the first count is  
15           greater than the non-anchor word count, then  
16           determining that the component is anchor-heavy.

1   15. A method for detecting objectionable navigation bars  
2   in a document, the method comprising:  
3       a) segmenting the document into components;  
4       b) for each of the components, determining whether or  
5       not the component is a navigation bar; and  
6       c) for each of the components that is determined to  
7       be a navigation bar, determining whether or not the  
8       navigation bar is disqualified from being classified  
9       as an objectionable navigation bar.

1   16. The method of claim 15 wherein the act of determining,  
2   for each of the components, whether or not the component is  
3   a navigation bar is based on a number of anchors in the  
4   component and a number of non-anchor words in the  
5   component.

1   17. The method of claim 15 wherein the act of determining  
2   whether or not the component is a navigation bar includes  
3       i) determining a number of anchors in the  
4       component,  
5       ii) determining a number of non-anchor words in  
6       the component, and

7           iii) if the number of anchors is greater than a  
8           predetermined threshold and if the number of  
9           anchors is greater than the number of non-anchor  
10          words, then determining that the component is a  
11          navigation bar.

1   18. The method of claim 15 wherein the act, for each of  
2   the components that is determined to be a navigation bar,  
3   of determining whether or not the navigation bar is  
4   disqualified from being classified as an objectionable  
5   navigation bar includes determining whether a  
6   disqualification condition, selected from a group of  
7   disqualification conditions consisting of (a) if the  
8   component has less than a predetermined number of anchors,  
9   (b) if the component has more than a predetermined  
10  percentage of words of the document, and (c) if the  
11  component is an element of a disqualified component and  
12  that disqualified component has only navigation bar  
13  elements, exists.

1   19. The method of claim 16 wherein the act, for each of  
2   the components that is determined to be a navigation bar,  
3   of determining whether or not the navigation bar is  
4   disqualified from being classified as an objectionable  
5   navigation bar includes determining whether a  
6   disqualification condition, selected from a group of  
7   disqualification conditions consisting of (a) if the  
8   component has less than a predetermined number of anchors,  
9   (b) if the component has more than a predetermined  
10  percentage of words of the document, and (c) if the  
11  component is an element of a disqualified component and

12 that disqualified component has only navigation bar  
13 elements, exists.

1 20. The method of claim 17 wherein the act, for each of  
2 the components that is determined to be a navigation bar,  
3 of determining whether or not the navigation bar is  
4 disqualified from being classified as an objectionable  
5 navigation bar includes determining whether a  
6 disqualification condition, selected from a group of  
7 disqualification conditions consisting of (a) if the  
8 component has less than a predetermined number of anchors,  
9 (b) if the component has more than a predetermined  
10 percentage of words of the document, and (c) if the  
11 component is an element of a disqualified component and  
12 that disqualified component has only navigation bar  
13 elements, exists.

1 21. A method for detecting objectionable navigation bars  
2 in a document, the method comprising:  
3 a) segmenting the document into components by  
4 generating a parse tree based on the document, wherein  
5 a first node corresponding to a first component is a  
6 child of a second node of a second component if the  
7 first component is included in the second component;  
8 b) for each of the nodes of the parse tree,  
9 determining whether or not the node corresponds to a  
10 navigation bar component; and  
11 c) for each of the nodes that is determined to  
12 correspond to a navigation bar, determining whether or  
13 not the navigation bar is disqualified from being  
14 classified as an objectionable navigation bar.

1 22. The method of claim 21 wherein the act, for each of  
 2 the nodes that is determined to correspond to a navigation  
 3 bar, of determining whether or not the navigation bar is  
 4 disqualified from being classified as an objectionable  
 5 navigation bar includes determining whether a  
 6 disqualification condition, selected from a group of  
 7 disqualification conditions consisting of (a) if the  
 8 component associated with the node has less than a  
 9 predetermined number of anchors, (b) if the component  
 10 associated with the node has more than a predetermined  
 11 percentage of words of the document, and (c) if the node  
 12 has a disqualified ancestor node and that all descendant  
 13 nodes of the disqualified ancestor node are associated with  
 14 navigation bar components, exists.

1 23. A machine-readable medium having machine executable  
 2 instructions thereon, wherein when the machine executable  
 3 instructions are executed on a machine, the machine:  
 4 a) segments the document into components; and  
 5 b) for each of the components, determines whether or  
 6 not the component is anchor-heavy, wherein if the  
 7 component is anchor-heavy, it is determined to be a  
 8 navigation bar.

1 24. A machine-readable medium having machine executable  
 2 instructions thereon, wherein when the machine executable  
 3 instructions are executed on a machine, the machine:  
 4 a) segments the document into components;  
 5 b) for each of the components, determines whether or  
 6 not the component is a navigation bar; and  
 7 c) for each of the components that is determined to  
 8 be a navigation bar, determines whether or not the

9 navigation bar is disqualified from being classified  
10 as an objectionable navigation bar.

1 25. An apparatus for detecting navigation bars in a  
2 document, the apparatus comprising:  
3 a) means for segmenting the document into components;  
4 and  
5 b) means for determining, for each of the components,  
6 whether or not the component is anchor-heavy, wherein  
7 if the component is anchor-heavy, it is determined to  
8 be a navigation bar.

1 26. An apparatus for detecting objectionable navigation  
2 bars in a document, the apparatus comprising:  
3 a) means for segmenting the document into components;  
4 b) means for determining, for each of the components,  
5 whether or not the component is a navigation bar; and  
6 c) means for determining, for each of the components  
7 that is determined to be a navigation bar, whether or  
8 not the navigation bar is disqualified from being  
9 classified as an objectionable navigation bar.